

High-Throughput Analysis of Clinical Flow Cytometry Data by Automated Gating

Hunjoong Lee, Yongliang Sun, Lisa Patti-Diaz, Michael Hedrick and Anka G Ehrhardt

Clinical Flow Cytometry, Translational Medicine, Bristol-Myers Squibb, Pennington, NJ, USA.

Bioinformatics and Biology Insights
Volume 13: 1–9
© The Author(s) 2019
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/1177932219838851



ABSTRACT: Advancements in flow cytometers with capability to measure 15 or more parameters have enabled us to characterize cell populations at unprecedented levels of detail. Beyond discovery research, there is now a growing demand to dive deeper into evaluating the immune response in clinical trials for immune modulating compounds. However, for high-volume, complex flow cytometry data generated in clinical trials, conventional manual gating remains the standard of practice. Traditional manual gating is resource intense and becomes a bottleneck and an impractical method to complete high volumes of flow cytometry data analysis. Current efforts to automate “manual gating” have shown that computational algorithms can facilitate the analysis of daunting multi-parameter data; however, a greater degree of precision in comparison with traditional manual gating is needed for wide-scale adoption of automated gating methods. In an effort to more closely follow the manual gating process, our automated gating pipeline was created to include negative controls (Fluorescence Minus One [FMO]) to enhance the reliability of gate placement. We demonstrate that use of an automated pipeline, heavily relying on FMO controls for population discrimination, can analyze multi-parameter, large-scale clinical datasets with comparable precision and accuracy to traditional manual gating.

KEYWORDS: Automated gating, flow cytometry, data analysis, drug-development

RECEIVED: February 8, 2019. **ACCEPTED:** February 13, 2019.

TYPE: Technical Advances

FUNDING: The author(s) received no financial support for the research, authorship, and/or publication of this article.

DECLARATION OF CONFLICTING INTERESTS: The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

CORRESPONDING AUTHOR: Hunjoong Lee, Clinical Flow Cytometry, Translational Medicine, Bristol-Myers Squibb, Pennington, NJ 08534, USA.
Email: Hunjoong.lee@bms.com

Introduction

In the past two decades, advancements in cytometry technologies give us unprecedented opportunity to characterize cell populations by surface and intracellular protein markers at the single-cell level. As flow cytometry continues to rapidly expand in system biology and medicine, it plays a role as a significant tool at all levels of drug discovery, translational, and clinical medicine.¹

Considering the increasing dimensionality of datasets, traditional manual gating of multi-parameter flow cytometry data is tedious and resource-demanding. In addition, manual gating can be subjective and time-prohibitive for a large amount of high-throughput data that is characteristic in a clinical setting.^{2,3}

In response to these issues, FlowCAP projects took the initiative to advance the development of computational methods for identification of cell populations and sample classification. A number of computational methods have been proposed to tackle these challenges and show that the level of computational analysis is matured for reliable use in data analysis.⁴

A series of software packages for automatic analysis implemented with different algorithms such as dimensionality reduction,^{5,6} mixture model-based clustering,⁷⁻⁹ artificial neural network,¹⁰ density-based clustering,^{11,12} and various novel algorithms¹²⁻¹⁴ have been published and expected to be widely used in the flow cytometry research community. For example, viSNE maps high-dimensional, single-cell data to lower dimension, typically two-dimensional (2D) plots, with the maintenance of data integrity. Non-Gaussian models such as

t -distribution and generalized t -distribution were adopted in flowClust and FLAME (FLow analysis with Automated Multivariate Estimation) to cluster cell populations automatically. FlowSOM equipped with a self-organizing map algorithm demonstrated superior performance with faster runtimes on test dataset compared with several methods.¹⁵ The current trend in data analysis software is shifted to automated population identification from multidimensional data using machine learning and statistical models.¹⁶⁻²² However, for clinical diagnostics, manual gating analysis is widely preferred and there are few published reports that apply automated gating on clinical data in large-scale studies. Gating rare cell populations with low cell counts is heavily dependent on Fluorescence Minus One (FMO) controls which contain all the fluorophores except for the one of negative controls.

In this article, we describe an automated gating pipeline, built using a number of open-source packages to extract all negative controls from FMO controls and apply these to fully stained samples to mimic manual gating process. We have analyzed 1698 samples of a T cell effector/memory panel and 1908 samples of regulatory T cell panel and evaluated the performance of automated gating analysis by comparison to standard manual gating analysis. Our primary goal is not to propose new clustering algorithms or novel automated gating algorithms. Instead, we demonstrate that the fast and reproducible automatic gating analysis of multi-parameter flow cytometry data can be reliably accomplished by open-source packages with additional monitoring steps in a workflow.



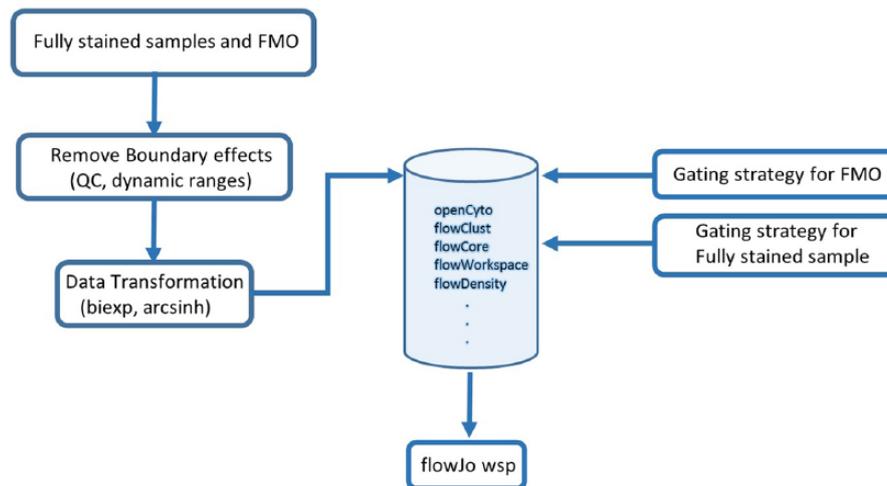


Figure 1. Schematic diagram of automated gating pipeline. The gating pipeline was composed of open-source packages with some customized modifications. Two gating templates, one for FMO and the other for fully stained sample, were prepared for hierarchical gating. Gating results including tables and PDF layout were repackaged as a FlowJo wsp file.

Materials and Methods

Panel information, sample staining, and acquisition

Two 12-color panels were developed and validated for clinical use and evaluated in this study: A T effector/memory panel including CD45, CD3, CD4, CD8, CD27, CD45RA, CCR7, PD1, HLA-DR, CTLA-4, CD38, and Ki67 and a regulatory T cell including CD45, CD3, CD4, CD25, CD127, CCR4, CD45RA, HLA-DR, CD39, ICOS, PD1, and CTLA-4. Fluorescence Minus One controls were set up for continuous expression markers such as PD1, HLA-DR, CTLA-4, and ICOS. Blood samples were collected in Streck Cyto-Chex BCT tubes, then stained for the above two panels, as described in our previous publication.²³ A Beckman Coulter's CytoFLEX S flow cytometer was used to acquire data from stained samples. The performance of each instrument is regularly monitored by beads to minimize the technical variation of samples. We collected each fully stained sample with more than 10 FMO samples in a same well to prevent technical variations.

Central manual gating

Manual gating was carried out with FlowJo version 10.2. For the T effector/memory panel, first, CD45⁺ leucocytes were gated based on CD45⁺ area versus side scatter area. After cell doublets were excluded using forward scatter area versus forward scatter width, CD3⁺ cells were gated, and then divided into CD3⁺CD4⁺ and CD3⁺CD8⁺ cells. For either CD3⁺CD4⁺ or CD3⁺CD8⁺ cells, CCR7 and CD45RA were used to further divide them into four subsets: naïve (N; CD45RA⁺CCR7⁺), central memory (CM; CD45RA⁻CCR7⁺), effector memory (EM; CD45RA⁻CCR7⁻), and terminally differentiated effector memory CD45RA⁺ (TEMRA; CD45RA⁺CCR7⁻). Continuous expression markers such as PD1, HLA-DR, and CTLA-4 were gated

based on 0.5% of their respective FMOs for each subset. For the regulatory T cell panel, similar approach was used to define CD4⁺ cells. Then, Treg populations were gated as CD25^{high}CD127^{low} from CD4⁺ cells. CD39, HLA-DR, CCR4, and CD45RA were used to further define the Treg populations. Again, “0.5% rule” was applied when continuous expression markers PD1, CTLA4, and ICOS were gated. All the defined reportable results were exported to comma-separated values (csv) files by FlowJo. All personnel performing manual gating were trained to follow standard guidelines and met appropriate requirements to be considered qualified. In addition, all manual gating was reviewed by an additional team member to verify the results.

Automated gating

The automated gating pipeline was built with multiple open-source packages and in-house post-processing modules (Figure 1).

Flow Cytometry Standard (FCS) files were transformed by bi-exponential function and compensated by a spillover matrix by an R package, flowCore.

Pre-defined hierarchical gating strategies were implemented in OpenCyto gating templates²⁴ where gating populations, gating methods, and parameters were declared. Generally, one-dimensional (1D) gating for FMO controls was done by the estimation of probability density function, followed by determining a cut-off point based on slopes. Those cut-off points for 1D gating from FMO controls were transferred to fully stained samples for population gating. To adopt the “0.5% rule” from our central manual gating, “adjust” and “tolerance” parameters of density function (R base) were tuned.

Two-dimensional gating for CD3⁺ populations in both T effector/memory panel and the regulatory T cell panel and for Treg populations in the regulatory T cell panel was carried out by flowClust.⁹

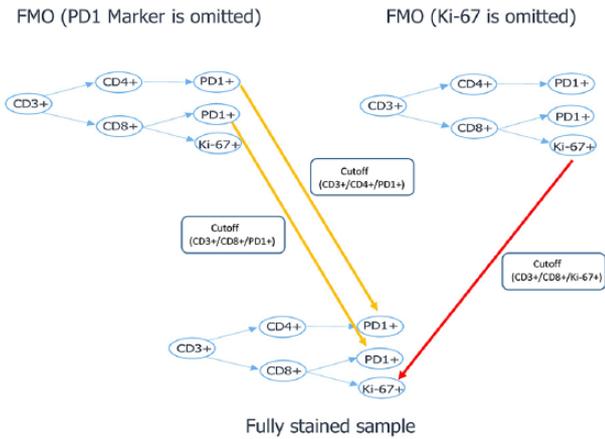


Figure 2. Scheme for applying FMO controls to a fully stained sample. For example, the cut-off of negative control for “CD3+/CD4+/PD1+” (fully stained sample) is obtained from “CD3+/CD4+/PD1+” population in FMO without PD1 marker.

flowClust was used to gate CD3+ and Treg populations with pre-calculated parameters such as “number of clusters,” “mean vector of centroid,” “covariance matrix,” and “degree of freedom.” One set of pre-calculated parameters for Treg and two sets of parameters for CD3+ populations were obtained.

Two gating templates for OpenCyto were made for each panel, one for FMO controls and the other for fully stained samples. The gating templates for FMO controls were used to extract negative controls for populations, and those negative controls were transferred to gate corresponding populations in fully stained samples (Figure 2). The cut-off points of the negative controls were determined by “tolerance” parameter applied on the slopes of probability density functions. The quality control of automated gating results was done by applying two filters. Since most populations were gated based on FMO controls, automated gating results without negative controls were flagged as a “failure.”

The second filter was applied to monitor whether target populations were correctly identified by clustering, for example, clustering with pre-calculated parameters incorrectly identified “CD3+” populations when the coordinates of the “CD3+” populations were significantly different from pre-calculated cluster centroid coordinates (Supplemental Figure S1). “CD3+” populations were incorrectly identified in 10.7% of the T effector/memory panel (182 out of 1698) and 6.8% of the regulatory T cell panel (129 out of 1908). Those CD3+ populations were re-gated with another set of pre-calculated parameters.

The automated gating results were reformatted into FlowJo (version 10.2) workspace (wsp) files using Perl XML module. Additional statistics tables and PDF layouts were also included in wsp file by editing FlowJo wsp files with Perl XML module.

Data and statistical analysis

The coefficient of variation (CV) of a cell population was calculated to show the robustness of gating methods. The R package, boot, (version 3.2.2) was used to perform bootstrapping

procedures with replications of 10000 to calculate the standard error of CV for automated gating and manual gating results. The average cell event of a population was calculated from automated gating analysis.

Correlation analysis of frequencies of populations from automated gating compared with manual gating was performed by an R package, stats. We analyzed the fold changes of cell populations in time-course data obtained from manual gating and automatic gating. Fold changes of cell populations compared with baseline levels were analyzed. Data with more than three time-points were used to calculate the similarity by cosine similarity scores (R package, lsa).

Results and Discussions

Automated gating strategy

To emulate manual gating process, the automated gating pipeline addresses three major points: (a) how to emulate bivariate manual gating of populations on two markers (eg Treg populations and CD3+ populations); (b) how to determine the cut-off points of negative controls from FMO controls and apply them to fully stained samples; and (c) data portability and visualization.

Previously published methods such as OpenCyto,²⁴ flowClust,⁹ flowDensity,¹⁶ and FlowSOM¹⁰ were tested to optimize the overall performance of the pipeline. We used OpenCyto as a main template which allowed users to define hierarchical gates with the selection of gating functions. One-dimensional (1D) gates were performed by gating functions provided by OpenCyto such as mindenisty and tailgate (Quad gates were a combination of two 1D gates). Clustering by FlowSOM based on a self-organizing map could be used to mimic manual gating, but did not give exact populations defined by pre-defined hierarchical gates, which hindered the direct comparison with the manual gating analysis.

The first point was addressed by flowClust (Figure 3). flowClust employed a *t*-mixture model-based clustering approach and needed the number of clusters to initiate the EM algorithm and returned the clusters with centroids and covariance matrix. In general, mixture model-based clustering did not scale well to large datasets because the EM algorithm was computationally expensive. As expected, automatic clustering by flowClust was a bottleneck step to process large number of cell events. To process our gating strategies for two T cell panels, we needed to run flowClust multiple times for around 10 FMO control FCS files. Furthermore, flowClust often incorrectly detected target populations that were not clearly separated from the rest of populations, which entailed fine-tuning of flowClust parameters (Figure 3). Therefore, we pre-calculated the parameters for multivariate *t*-distribution with “training sets” iteratively, for example, CD3+ populations and Treg populations, and provided those parameters to flowClust for facilitating the EM algorithm. Generally, for CD3+ populations in the 2D dimension of SSC and CD3, two sets of parameters were obtained. There are outstanding unsupervised

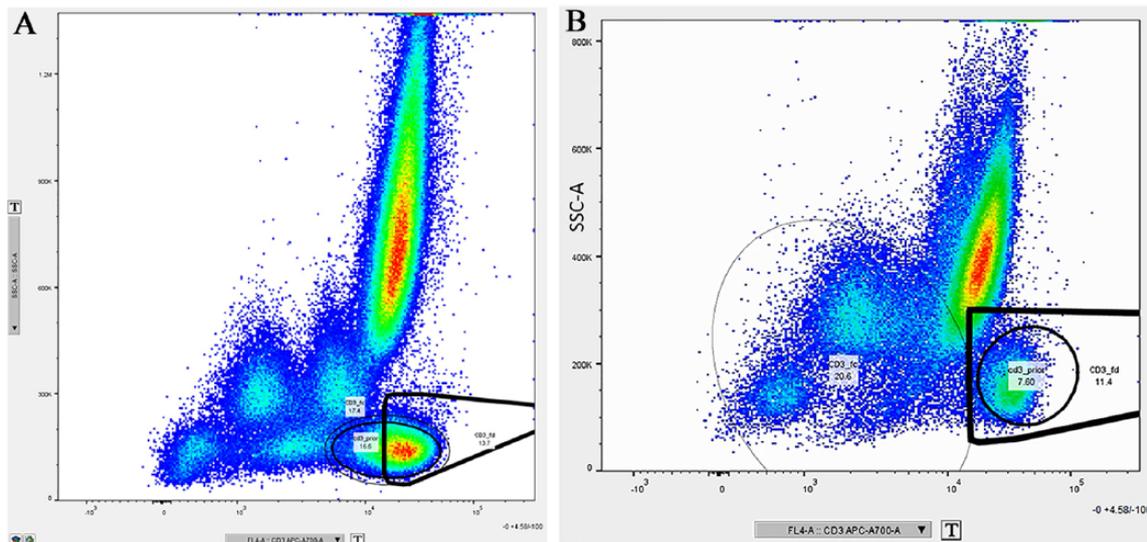


Figure 3. CD3+ gate by flowClust (normal line), flowClust with prior information (heavy line), and flowDensity (thick line): Three different methods gate CD3+ correctly (A) and clustering with prior information is better for weak CD3+ populations than flowClust and flowDensity (B).

methods for data clustering published^{10,12} and recently constrained clustering is also available,²² but it is unlikely that a single method can fit the diversity of samples.

As for the second point, the cut-off points of negative controls from FMO controls were determined by the slopes of probability density functions and transferred to the fully stained samples. It is generally recommended to adopt the negative controls for gating of “rare populations” which are loosely defined by low frequencies of populations (as % of parent <5).²¹ Rare populations are often of biological importance, and the use of negative controls is critical to identify those populations with accuracy compared with the manual gating analysis. To extract all negative controls for cell populations, the automated gating was applied to all FMO controls to determine cut-off points for populations, and those cut-off points were saved in temporary holders and subsequently applied to fully stained samples (Figure 3; Supplemental Figure S2). As probability density functions were estimated by kernel density function, it was natural that a smoothing parameter needed to be tuned for the rare populations. We found that most of rare populations could be estimated well enough by the “adjust” parameter in R density function. To determine the cut-off point, the “tolerance” parameter needed to be set appropriately. For example, OpenCyto sets the default “tolerance” at 0.01 to gate a shoulder peak population from a main peak population. The manual gating analysis using FMO controls was carried out. To reduce the technical variability from manual gating, “0.5% rule” is applied for negative gate placement. In the automated gating analysis, we use the “tolerance” parameter to emulate the manual gating rule of “0.5%.” Different functions such as quantileGate function (OpenCyto) could be used to determine the cut-off points, and the gate transfer from FMO controls to fully stained samples could be also done by other packages such as flowDensity.¹⁶

The last important point of the automated gating pipeline was an effective method of data portability and the visualization of automated gating analysis. The increased complexity of data required the visualization of the automated gating analysis for data interpretation and for a better communication with experimental scientists. Even though CD3+ populations were correctly identified by two sets of pre-calculated parameters, it is critical to provide visualization tools due to the diversity of samples. We generated FlowJo wsp files to contain automatic gating results of FMO controls and fully stained samples along with additional tables and PDF layout using Perl XML module to facilitate quick visual examination of results (Figure 4). Currently, there is also an open-source package (CytoML) providing methods to export automatic analysis to FlowJo wsp files, but more customized steps might be necessary for individual study.²⁵

Post-process of a large-scale automated gating

Quality control in a large-scale automated gating is crucial in both manual analysis and automatic analysis because of high diversity of samples.

No single model can fit all data with high variation, and it was not feasible to check thousands of automated gating results visually for quality control. Therefore, setting QC filters was crucial in the automated gating analysis.

In our automatic pipeline, the transference of FMO cut-off points to fully stained samples for 1D gates and identification of target populations by flowClust for 2D gates were often main causes of the discrepancy from the manual gating analysis. Therefore, we applied two filters to flag a “failed” analysis: (a) automated gating failed to determine the cut-off points for populations from FMO controls mainly due to the poorly defined kernel density estimate of cell populations and (b)

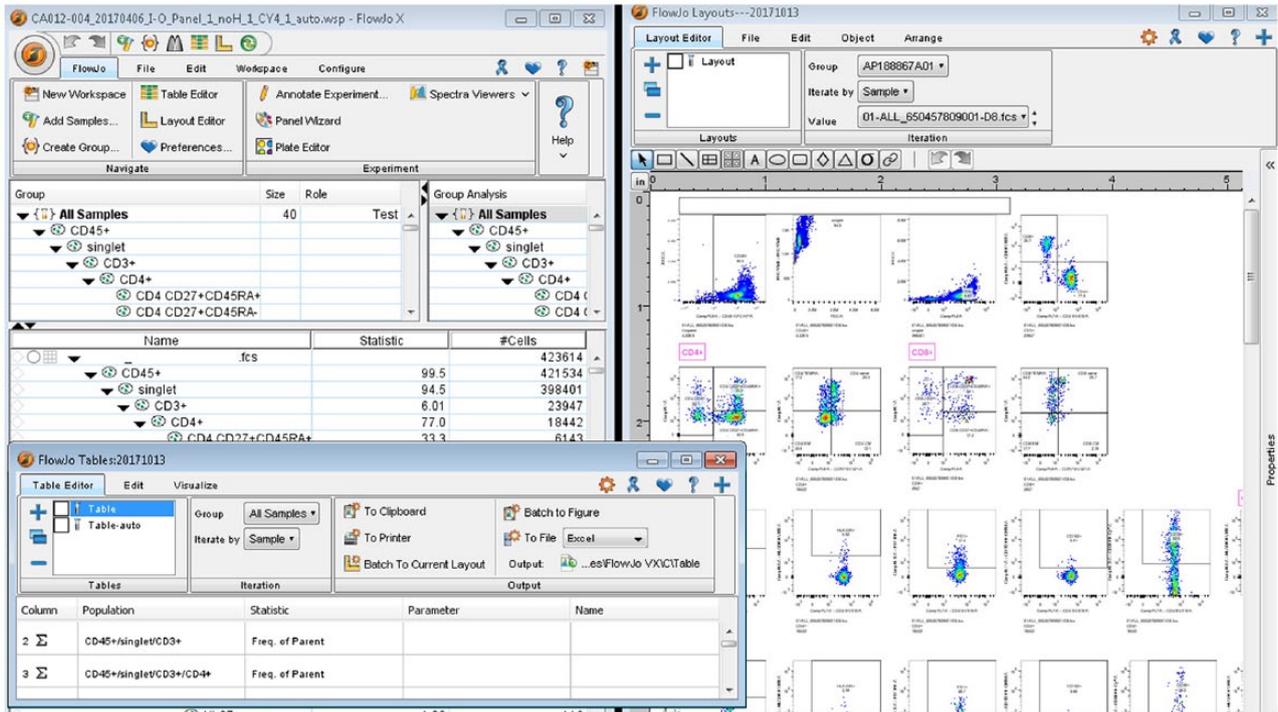


Figure 4. FlowJo workspace file generated by automated gating pipeline. Automated gating results were packaged into a FlowJo wsp file with all information including statistics, gating hierarchy, tables, and PDF layout.

automated gating incorrectly identified target cell populations due to the significant coordinate difference of target populations from pre-calculated parameters. For example, the detection of CD3⁺ populations in our panels was dependent on the side scatter channel value and the CD3 fluorescence value (coordinates). When there were samples with abnormal coordinates due to the diversity of samples, the failed automated gating could be detected by monitoring the frequency of CD3⁺ populations (Supplemental Figure S1).

In our panels, around 10% of CD3⁺ populations with high diversity of samples were completely missed and detected by the filter (Supplemental Figure S1). Those failed populations were gated correctly by applying flowClust with other prior parameters.

The design of hierarchical gating strategy was important to apply the automated gating to large-scale immunology human study. For example, gating strategy where populations with high variability were gated by 1D gate rather than 2D gate and the selection of appropriate bright fluorophores for rare populations would make the automated gating analysis more robust and reliable.

Performance evaluation of automated gating

The robustness of automated gating was evaluated by comparing the CV of manual gating and automated gating of baseline clinical samples (Figure 5). Due to different ways for negative gate placement of manual gating (“0.5% rule” based on cell counts) and automated gating (“adjust” and “tolerance”), the CV in

Figure 5 is considered as indirect comparison of two gating results, but the CVs of most populations from the automated gating analysis were comparable to those from the manual gating analysis. Generally, the CVs from automated gating were similar to those of manual gating for most populations such as CD3⁺, CD4⁺, and CD8⁺ as previously reported elsewhere.^{26,27} The rare populations typically with low cell counts showed much higher CV in the automated gating analysis, implying that either probability density functions were not well defined or cut-off points were determined differently from those by the manual gating analysis. Memory T cells such as CD4 CM/EM and CD8 CM/EM were known as poorly resolved cell populations and were likely to be subjected to individual interpretation.²⁶ In our study, automated gating for those memory T cells is comparable to central gating (Supplemental Figure S3). Another notable observation was made with Treg populations. The Treg populations were labeled as reliable populations according to the Human ImmunoPhenotyping Consortium (HIPIC) panel,²⁷ but classified as “poorly resolved” with high CVs due to the lack of clear markers and FOXP3 was suggested.²⁶ In our case, the CVs of Treg populations from manual gating and clustering by pre-calculated parameters were comparable as reported in HIPIC.

It was known that both automated and manual gating of most cell populations with high cell events were in good agreement. In our analysis, automated gating was also in good agreement with manual gating for high event cell populations (Figure 6; Supplemental Figure S3). The poorly resolved populations or rare populations with low cell events often result in subjective and non-reproducible gating.

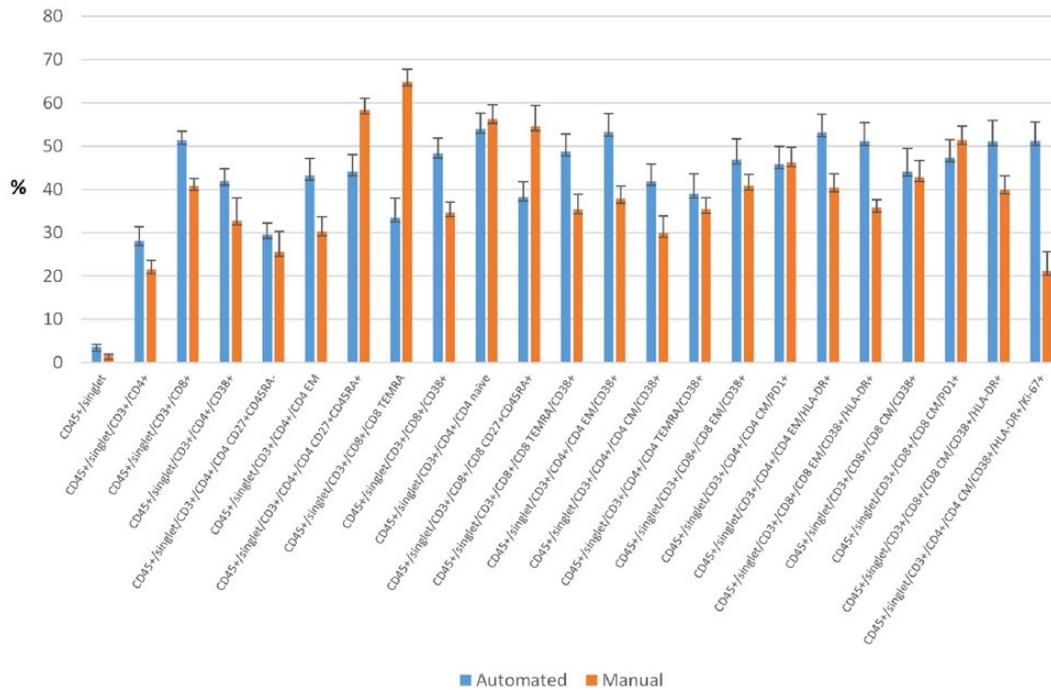


Figure 5. Coefficient of variation from automated and manual gating. X-axis represents cell populations sorted by CV of automated gating. Coefficients and standard errors of automated gating and manual gating are shown in cyan bar and orange bar, respectively. The populations are sorted by abundance. A total of 72 clinical baseline samples were used.

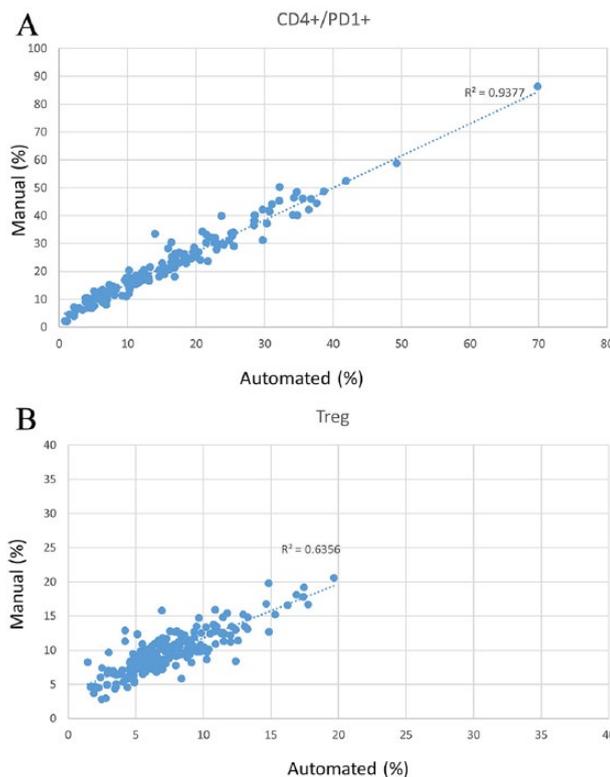


Figure 6. Linear regression plots of population frequencies from automated gating (x-axis) and manual gating for (A) CD4+/PD1+ which were gated based on FMO (149 samples) and (B) Tregs (CD25, CD127) (211 samples) are shown.

As authors in flowLearn²¹ illustrated an analysis bias per center of the manual gating analysis on FlowCAP data, the discrepancy between automated gating results and manual

gating results was expected to be persistent especially for poorly resolved populations and rare populations. Thus, it was reasonable to compare alternative metrics, such as in our case, fold

changes of populations, to evaluate the automated gating analysis.

To determine whether automated gating and manual gating were comparable in terms of data interpretation, we analyzed the time-course data of multiple data points (at least 3) from 44 subjects. The fold changes of cell populations to baseline level from manual gating and automated gating were compared by cosine similarity scores (Figure 7). Even poorly resolved populations or rare populations typically with low cell counts showed high similarity scores in fold changes (Figure 7B), indicating that manual gating and automated gating could draw similar interpretation of gated data.

Poorly resolved populations such as “CD8+/CD8 TEMRA” and “CD8+/CD8 CD27+CD45RA+” (Supplemental Figure S4) showed higher CV in manual gating (Figure 5), clearly indicating that automated gating could help reducing subjective bias of manual gating. In general, as the average cell counts for populations decreased, CV showed a gradual increase as expected.

The measurable number of markers increases with recent advance in instrumentation, and pre-defined hierarchical gating will play an important role in clinical trials. There are a number of elegant unsupervised algorithms and semi-supervised methods to analyze multi-parameter cytometry data, but we found that it was unlikely to fit all biological samples with high variations (Figure 3). The gating discrepancies of rare populations, especially gated by FMO controls, could stem from either incorrect density estimation of cell events in automated gating or inconsistent application of “0.5% rule” in manual gating. Those discrepancies can be reduced by fine-tuning parameters with the help of manual gating operators (Supplemental Figure S2).

In our study, relatively simple fine-tuning of parameters such as “adjust” and “tolerance” and parameters for mixture model-based clustering such as “centroids” and “number of clusters” allowed us to analyze a large number of clinical data with precision comparable to manual gating analysis. However, based on our high-throughput analysis of clinical data using the automated gating pipeline, we suggest that it should be essential to have additional steps to detect outliers stemmed from the diversity of samples, for example, monitoring certain populations, and also provide visualization tools for quick manual examination.

Conclusions

It is impossible to monitor changes in immune profiles of subjects in large-scale ongoing clinical trials with traditional manual data analysis, which necessitated the development of a robust alternative computational method. Multi-parameter cytometry becomes an essential technique for characterizing individual immune traits, and automated gating will be essential to handle large-scale datasets with comparable precision and accuracy to the manual gating with reproducibility. In

addition, numerous reports on human immune trait variations have been published, suggesting non-inheritable factors such as the shared environmental factors and microbes were accountable for immune cell profiles to a larger extent than we expected.^{28–30} Systematic discrepancies for populations, especially gated by FMO controls, between the manual gating analysis and the automated gating analysis could be reduced by tuning “adjust” or “tolerance” parameters. As authors in flow-Learn paper clearly showed the variability of the manual gating analysis on FlowCAP data, we believed that the reproduction of manual gating analysis was not an ultimate metric for evaluation of automated gating analysis. The automated gating analysis could deliver robust, reproducible, and faster analysis than manual gating analysis did. However, fine-tuning of parameters and selection of gate functions were essential due to high diversity of samples which also showed the importance of proper quality check for the automated gating analysis.

In conclusion, we built automated gating pipeline incorporating FMO control gating and demonstrated the feasibility of robust automated gating to process large-scale datasets with reproducibility in comparison with manual gating.

Acknowledgements

The authors thank Bristol-Myers Squibb Clinical Flow Cytometry team for processing clinical samples, and Syngene International Limited, Biocon Bristol-Myers Squibb R&D Center in India for manual gating.

Author Contributions

This study was designed by HL, YS, MH, and AE. The pipeline was implemented and evaluated by HL, YS, and LPD wrote the paper with feedback from MH and AGE.

Supplemental material

Supplemental material for this article is available online.

REFERENCES

- Mair F, Hartmann FJ, Mrdjen D, Tosevski V, Krieg C, Becher B. The end of gating? An introduction to automated analysis of high dimensional cytometry data. *Eur J Immunol*. 2016;46:34–43.
- Maecker HT, McCoy JP, Nussenblatt R. Standardizing immunophenotyping for the Human Immunology Project. *Nat Rev Immunol*. 2012;12:191–200.
- Saeys Y, Van Gassen S, Lambrecht BN. Computational flow cytometry: helping to make sense of high-dimensional immunology data. *Nat Rev Immunol*. 2016;16:449–462.
- Aghaeepour N, Finak G; FlowCAP Consortium; DREAM Consortium, Hoos H, et al. Critical assessment of automated flow cytometry data analysis techniques. *Nat Methods*. 2013;10:228–238.
- Van der Maaten L, Hinton G. Visualizing data using t-SNE. *J Mach Learn Res*. 2008;9:2579–2605.
- Amir AD, Davis KL, Tadmor MD, et al. viSNE enables visualization of high dimensional single-cell data and reveals phenotypic heterogeneity of leukemia. *Nat Biotechnol*. 2013;31:545–552.
- Pyne S, Hu X, Wang K, et al. Automated high-dimensional flow cytometric data analysis. *Proc Natl Acad Sci U S A*. 2009;106:8519–8524.
- Sörensen T, Baumgart S, Durek P, Grützkau A, Häupl T. immunoClust—an automated analysis pipeline for the identification of immunophenotypic signatures in high-dimensional cytometric datasets. *Cytometry A*. 2015;87:603–615.
- Lo K, Hahne F, Brinkman RR, Gottardo R. flowClust: a bioconductor package for automated gating of flow cytometry data. *BMC Bioinformatics*. 2009;10:145.

10. Van Gassen S, Callebaut B, Van Helden MJ, et al. FlowSOM: using self-organizing maps for visualization and interpretation of cytometry data. *Cytometry A*. 2015;87:636–645.
11. Qian Y, Wei C, Eun-Hyung Lee F, et al. Elucidation of seventeen human peripheral blood B-cell subsets and quantification of the tetanus response using a density-based method for the automated identification of cell populations in multidimensional flow cytometry data. *Cytometry B Clin Cytom*. 2010;78: S69–S82.
12. Qiu P, Simonds EF, Bendall SC, et al. Extracting a cellular hierarchy from high-dimensional cytometry data with SPADE. *Nat Biotechnol*. 2011;29:886–891.
13. Levine JH, Simonds EF, Bendall SC, et al. Data-driven phenotypic dissection of AML reveals progenitor-like cells that correlate with prognosis. *Cell*. 2015;162:184–197.
14. Mosmann TR, Naim I, Rebhahn J, et al. SWIFT—scalable clustering for automated identification of rare cell populations in large, high-dimensional flow cytometry datasets, part 2: biological evaluation. *Cytometry A*. 2014;85:422–433.
15. Weber LM, Robinson MD. Comparison of clustering methods for high-dimensional single-cell flow and mass cytometry data. *Cytometry A*. 2016;89: 1084–1096.
16. Malek M, Taghiyar M, Chong L, Finak G, Gottardo R, Brinkman RR. flowDensity: reproducing manual gating of flow cytometry data by automated density-based cell population identification. *Bioinformatics*. 2015;31:606–607.
17. Van Gassen S, Vens C, Dhaene T, Lambrecht BN, Saeys Y. FloReMi: flow density survival regression using minimal feature redundancy. *Cytometry A*. 2016;89:22–29.
18. Bruggner RV, Bodenmiller B, Dill DL, Tibshirani RJ, Nolan GP. Automated identification of stratifying signatures in cellular subpopulations. *Proc Natl Acad Sci U S A*. 2014;111:E2770–E2777.
19. Lin L, Finak G, Ushey K, et al. COMPASS identifies T-cell subsets correlated with clinical outcomes. *Nat Biotechnol*. 2015;33:610–616.
20. Arvaniti E, Claassen M. Sensitive detection of rare disease-associated cell subsets via representation learning. *Nat Commun*. 2017;8:14825.
21. Lux M, Brinkman RR, Chauve C, et al. flowLearn: fast and precise identification and quality checking of cell populations in flow cytometry. *Bioinformatics*. 2018;34:2245–2253.
22. Lee A, Chang I, Burel JG, et al. DAFI: a directed recursive data filtering and clustering approach for improving and interpreting data clustering identification of cell populations from polychromatic flow cytometry data. *Cytometry A*. 2018;93:597–610.
23. Sun Y, Yang K, Bridal T, Ehrhardt AG. Robust Ki67 detection in human blood by flow cytometry for clinical studies. *Bioanalysis*. 2016;8:2399–2413.
24. Finak G, Frelinger J, Jiang W, et al. OpenCyto: an open source infrastructure for scalable, robust, reproducible, and automated, end-to-end flow cytometry data analysis. *PLoS Comput Biol*. 2014;10:e1003806.
25. Jiang M. *CytoML: GatingML Interface for OpenCyto. R Package Version 1.6.4*; 2016.
26. Burel JG, Qian Y, Lindestam Arlehamn C, et al. An integrated workflow to assess technical and biological variability of cell population frequencies in human peripheral blood by flow cytometry. *J Immunol*. 2017;198:1748–1758.
27. Finak G, Langweiler M, Jaimes M, et al. Standardizing flow cytometry immunophenotyping analysis from the Human Immunophenotyping Consortium. *Sci Rep*. 2016;6:20686.
28. Roederer M, Quaye L, Mangino M, et al. The genetic architecture of the human immune system: a bioresource for autoimmunity and disease pathogenesis. *Cell*. 2015;161:387–403.
29. Brodin P, Jovic V, Gao T, et al. Variation in the human immune system is largely driven by non-heritable influences. *Cell*. 2015;160:37–47.
30. Kaczorowski KJ, Shekhar K, Nkulikiyimfura D, et al. Continuous immunotypes describe human immune variation and predict diverse response. *Proc Natl Acad Sci U S A*. 2017;114:E6097–E6106.